

Revista Eletrônica de Sistemas de Informação

ISSN 1677-3071

v. 13, n. 3

set-dez 2014

doi:10.21529/RESI.2014.1303

Sumário

Editorial

[Editorial](#)

Pietro Cunha Dolci

Ensino e pesquisa

[SISTEMAS DE INFORMAÇÃO – TEMAS DE PESQUISA ACADÊMICA NO BRASIL ENTRE 1994 E 2013](#)

Henrique Freitas, João Becker, Cristina Martens, Carla Marcolin

Foco na tecnologia

[CRITÉRIOS DE SELEÇÃO DE SGBD NOSQL EM ORGANIZAÇÕES BRASILEIRAS](#)

Alexandre Moraes de Souza, Edmir Parada Vasques Prado

Foco nas organizações

[A COMUNICAÇÃO NO ALINHAMENTO ESTRATÉGICO TI-NEGÓCIO: O PAPEL DAS CONVERSACÕES E SUAS COMPETÊNCIAS](#)

Luís Kalb Roses, Jean Carlos Borges Brito, Gentil José de Lucena Filho

[SOFTWARE LIVRE COMO FATOR DE INOVAÇÃO PARA O SETOR DE TIC](#)

Luciano Cunha de Sousa, Cleidson Nogueira Dias, Solange Alfinito

Pensata

[FUTEBOL E A PESQUISA EM SISTEMAS DE INFORMAÇÃO: AFIRMANDO A IDENTIDADE BRASILEIRA](#)

Maria Alexandra Cunha, Eduardo Henrique Diniz

Nominata de avaliadores

[Avaliadores ad hoc - 2014](#)



Este trabalho está licenciado sob uma [Licença Creative Commons Attribution 3.0](http://creativecommons.org/licenses/by/3.0/).

ISSN: 1677-3071

Esta revista é (e sempre foi) eletrônica para ajudar a proteger o meio ambiente, mas, caso deseje imprimir esse artigo, saiba que ele foi editorado com uma fonte mais ecológica, a *Eco Sans*, que gasta menos tinta.

This journal is (and has always been) electronic in order to be more environmentally friendly. Now, it is desktop edited in a single column to be easier to read on the screen. However, if you wish to print this paper, be aware that it uses Eco Sans, a printing font that reduces the amount of required ink.

CRITÉRIOS DE SELEÇÃO DE SGBD NOSQL EM ORGANIZAÇÕES BRASILEIRAS

NOSQL DBMS SELECTION CRITERIA IN BRAZILIAN ORGANIZATIONS

(artigo submetido em maio de 2014)

Alexandre Morais de Souza

Programa de Pós-graduação em Sistemas de Informação - Universidade de São Paulo (USP)
omegamasters@gmail.com

Edmir Parada Vasques Prado

Programa de Pós-graduação em Sistemas de Informação - Universidade de São Paulo (USP)
eprado@usp.br

ABSTRACT

The large volume of data generated by Web applications and the requirements of scalability and high degree of availability have contributed to the emergence of new models and technologies related to DBMS (Database Management Systems). In this environment of growing demand for data volume and performance, a new model of DBMS, called NoSQL arises. Within this context, this research aims to identify the main criteria that should be used for selecting a NoSQL DBMS in private organizations. We used a qualitative research based on the Delphi Technique. The application of this technique was carried out between January and March 2013, with the participation of 32 specialists. The research identified nine criteria that should be considered in the selection process of DBMS in private organizations.

Key-words: DBMS; NoSQL; Delphi technique

RESUMO

O grande volume de dados gerado por aplicações Web, juntamente com requisitos diferenciados, como a escalabilidade e o elevado grau de disponibilidade, têm contribuído para o surgimento de novos modelos e tecnologias com relação aos SGBDs (Sistemas Gerenciadores de Bancos de Dados). Neste ambiente de crescente demanda por volume de dados e desempenho, surge um novo modelo de SGBD, chamado NoSQL. Dentro desse contexto, esta pesquisa tem como objetivo identificar os principais critérios que devem ser usados para seleção de um SGBD NoSQL em organizações privadas. Utilizou-se uma pesquisa qualitativa baseada na Técnica Delphi. A aplicação dessa técnica foi realizada entre os meses de janeiro e março de 2013 com a participação de 32 especialistas. A pesquisa identificou nove critérios que devem ser considerados no processo de seleção de SGBD em organizações privadas.

Palavras-chave: SGBD; NoSQL; técnica Delphi



1 INTRODUÇÃO

O grande volume de dados gerado por aplicações Web, juntamente com requisitos diferenciados, como a escalabilidade e o elevado grau de disponibilidade, têm contribuído para o surgimento de novos modelos e tecnologias com relação aos SGBDs (Sistemas Gerenciadores de Bancos de Dados).

Neste ambiente de crescente demanda por volume de dados e desempenho, surge um novo modelo de SGBD, chamado NoSQL, ou seja, não apenas SQL (*Structured Query Language*). Este novo modelo foi proposto com o objetivo de atender ao gerenciamento de grandes volumes de dados semiestruturados ou não estruturados, que necessitam de alta disponibilidade e escalabilidade (LÓSCIO; OLIVEIRA; PONTES, 2011). A necessidade de um novo conceito em banco de dados surgiu como consequência da ineficiência de SGBDs relacionais em lidar com o atual volume de informações das aplicações Web. Os SGBDs relacionais foram propostos na década de 70, quando as aplicações lidavam apenas com dados estruturados, ou seja, dados que possuem uma estrutura fixa e bem definida. Além disso, o volume de dados gerado por tais aplicações é muito menor do que o volume gerado pelas aplicações atuais na Web.

A motivação deste trabalho se deve aos poucos estudos relacionados com processos de seleção de SGBDs NoSQL. Esta pesquisa tem a finalidade de ajudar a preencher essa lacuna. Dentro desse contexto, esta pesquisa tem como objetivo geral identificar os principais critérios que devem ser usados para seleção de um SGBD NoSQL em organizações privadas. Para consecução desse objetivo geral foram definidos três objetivos específicos: (1) identificar na literatura os critérios de seleção de SGBDs; (2) identificar com especialistas, a partir dos critérios encontrados na literatura, os principais critérios para seleção de SGBDs NoSQL; e (3) ordenar os critérios de seleção de SGBDs NoSQL conforme a importância atribuída pelos especialistas.

2 SISTEMAS GERENCIADORES DE BANCOS DE DADOS NOSQL

Os SGBDs NoSQL não fazem uso da álgebra relacional, nem possuem linguagem de consulta nativa como a linguagem SQL utilizada em SGBDs relacionais tradicionais, o que torna mais complexa a atividade de desenvolvimento. Normalmente os SGBDs NoSQL possuem estrutura simplificada, sem estrutura de relacionamentos, com suporte natural a replicação. São também conhecidos como banco de dados escaláveis e direcionados para aplicações Web, devido à característica inerente de ter bom desempenho com grande volume de dados (TIWARI, 2011).

Os SGBDs NoSQL apresentam algumas características fundamentais que os diferenciam dos tradicionais SGBDs relacionais, tornando-os adequados para armazenamento de grandes volumes de dados não estruturados ou semiestruturados. Segundo Lóscio, Oliveira e Pontes (2011) algumas destas características são:

a) escalabilidade horizontal. A ausência de controle de bloqueios é uma característica dos SGBDs NoSQL que torna esta tecnologia adequada para solucionar problemas de gerenciamento de volumes de dados que crescem exponencialmente, como os dados da Web 2.0.

b) ausência de esquema ou esquema flexível. Uma característica evidente dos SGBDs NoSQL é a ausência completa ou quase total do esquema que define a estrutura dos dados modelados. Esta ausência facilita tanto a escalabilidade quanto contribui para o aumento da disponibilidade. Em contrapartida, não há garantias da integridade dos dados, o que ocorre nos bancos de dados relacionais, devido à sua estrutura rígida.

c) replicação de forma nativa. Diminui o tempo gasto para recuperar informações. Existem duas abordagens principais para replicação: *master-slave* (mestre-escravo) e *multi-master*.

d) consistência eventual. A consistência nem sempre é mantida entre os diversos pontos de distribuição de dados. Esta característica tem como princípio o teorema CAP (*consistency, availability and partition tolerance*), que diz que, em um dado momento, só é possível garantir duas de três propriedades entre consistência, disponibilidade e tolerância à partição (YE; LI, 2011).

Lóscio, Oliveira e Pontes (2011) classificaram os principais tipos de modelos de dados NoSQL em: chave-valor, orientado a colunas, orientado a documentos e orientado a grafos, os quais são descritos nas subseções a seguir.

2.1 MODELO DE DADOS CHAVE-VALOR

O modelo chave-valor (*key-value*) é considerado simples e permite a visualização do banco de dados como uma grande tabela *hash*, que é uma estrutura de dados especial que associa chaves de pesquisa a valores. De maneira bem simples, o banco de dados é composto por um conjunto de chaves, as quais estão associadas a um único valor (TIWARI, 2011).

Essa estrutura de dados é extremamente popular porque proporciona grande eficiência. Um par chave-valor é um valor único que pode ser facilmente utilizado para acessar dados. Este modelo é de fácil implementação, permite que os dados sejam rapidamente acessados pela chave e contribui para aumentar a capacidade de acesso aos dados em sistemas que possuem alta escalabilidade (TIWARI, 2011).

Na Figura 1 há uma representação da estrutura chave-valor com um exemplo de aplicação para *e-commerce*. Este é um típico exemplo onde se pode combinar o uso do SGBD NoSQL orientado a chave-valor e o modelo relacional.

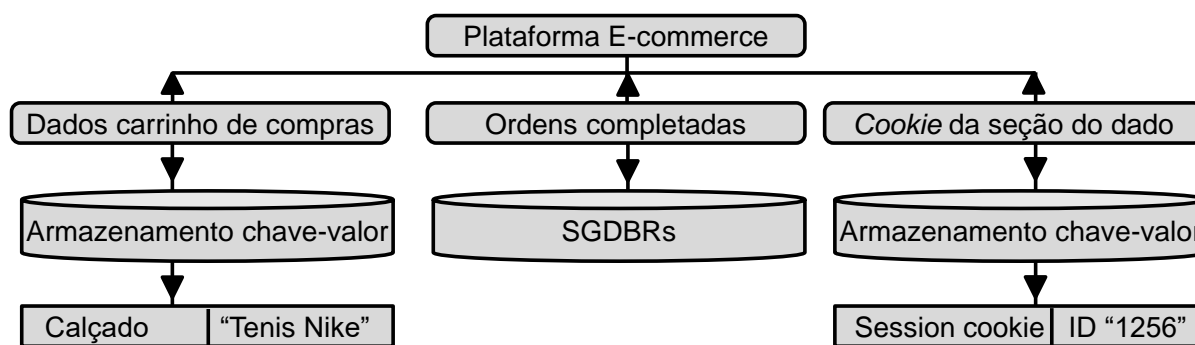


Figura 1. Banco de Dados NoSQL com abordagem Chave-Valor

Fonte: adaptado de Sadalage, Fowler e River (2013)

2.2 MODELO DE DADOS ORIENTADO A COLUNAS

O modelo orientado a colunas é mais complexo que o modelo chave-valor (POKORNY, 2011). Neste modelo os dados são indexados por uma tripla (linha, coluna e *timestamp*), onde linhas e colunas são identificadas por chaves e o *timestamp* permite diferenciar múltiplas versões de um mesmo dado. Outro conceito associado ao modelo é o de família de colunas (*column family*), que é usado com o intuito de agrupar colunas que armazenam o mesmo tipo de dados (POKORNY, 2011).

Segundo Tiwari (2011), os bancos orientados a colunas são os mais populares entre os bancos não relacionais. Embora os bancos de dados orientados a colunas existam de diferentes formas há alguns anos, eles foram popularizados para a comunidade de desenvolvedores com as publicações de artigos do Google© (GHEMAWAT; GOBIOFF; LEUNG, 2003; DEAN; GHEMAWAT, 2008; CHANG; DEAN; GHEMAWAT, 2008).

Apesar dos SGBDs orientados a colunas possuírem semelhanças com o modelo relacional, eles são diferentes. Não se pode aplicar o mesmo tipo de solução utilizado na forma relacional para um banco de dados orientado a colunas (RAHIEN, 2010). A Figura 2 apresenta um exemplo desse modelo.

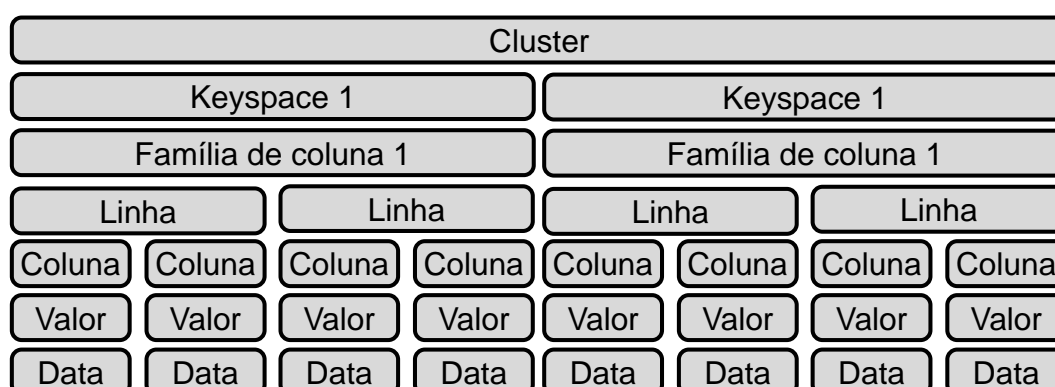


Figura 2. Modelo família de colunas

Fonte: adaptado de Sadalage, Fowler e River (2013)

2.3 MODELO DE DADOS ORIENTADO A DOCUMENTOS

O modelo orientado a documentos armazena coleções de documentos. No modelo chave-valor apenas uma única tabela *hash* é criada para todo o banco. No modelo orientado a documentos existe um conjunto de documentos e em cada documento há um conjunto de campos (chaves) e o valor deste campo (POKORNY, 2011).

As bases de dados orientadas a documentos possuem o foco em resolver problemas de flexibilidade causados por bancos de dados relacionais (TIWARI, 2011). Trata-se do esforço necessário para traduzir objetos de um SGBDR, o que se torna um empecilho para a produtividade do programador quando surge a necessidade de alterar alguma característica da modelagem de dados. Existem ferramentas, como *Hibernate*, para automatizar esse mapeamento, mas apenas resolvem parcialmente o problema (HARRISON, 2010). Para Redmond e Wilson (2012), a modelagem de dados em bancos orientados a documentos é mais flexível quanto a alterações. Para esses autores, nos bancos relacionais qualquer tipo de alteração no modelo de dados é custoso, pois os programas precisam ser modificados e compilados em conjunto com a alteração do modelo. Em bases de dados maiores, alteração no modelo envolve a propagação de alterações em vários outros bancos compartilhados. Nas bases de dados orientadas a documentos, uma aplicação pode modificar a estrutura do documento quanto quiser sem impacto para a aplicação, com o custo de riscos de inconsistências ou a existência de documentos obsoletos (HARRISON, 2010).

Na Figura 3 é apresentado o funcionamento da estrutura de *sharding*, onde os dados são divididos em vários *shards*, cada um contando com sua própria estrutura de replicação.

Segundo Harrison (2010), em uma base de dados orientada a documentos, o documento pode mapear quase que diretamente para linguagem de estrutura de classes. Para esse autor, isto faz com que a programação seja mais fácil, mas faz levantar questões de integridade de dados, uma vez que alguns itens de dados são quase inevitavelmente duplicados.

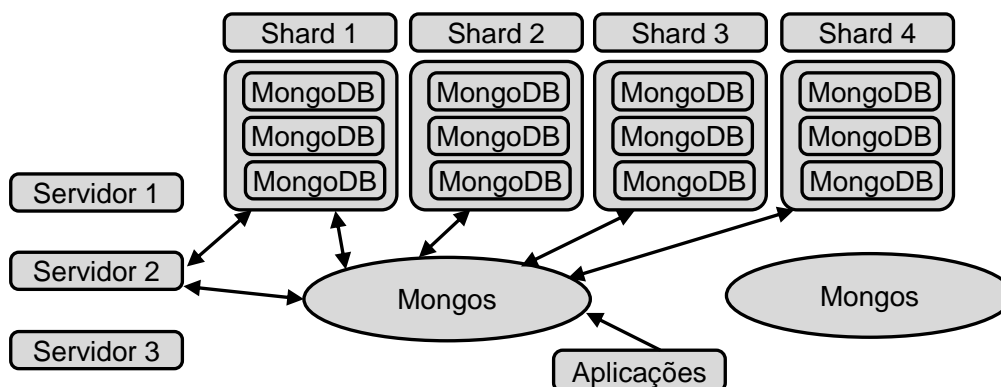


Figura 3. Estrutura de *Sharding*
Fonte: adaptado de Tiwari (2011)

2.4 MODELO DE DADOS ORIENTADO A GRAFOS

O modelo orientado a grafos, segundo Sadalage, Fowler e River (2013), possui três componentes básicos: os nós (vértices do grafo), os relacionamentos (as arestas) e as propriedades dos nós e relacionamentos. Para Pokorny (2011), este modelo permite o armazenamento das entidades e dos relacionamentos entre elas. Assim, pode-se armazenar o dado de uma forma e interpretá-lo de diferentes formas baseado no seu relacionamento. A Figura 4 ilustra um exemplo desse tipo de modelo.

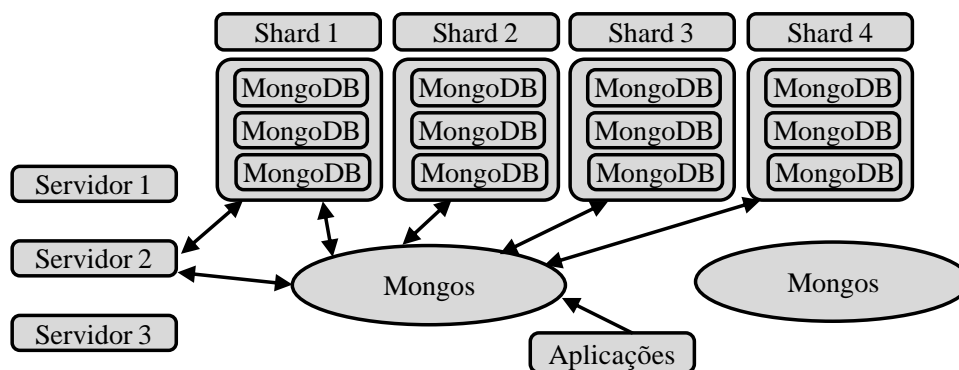


Figura 4. Banco de dados NoSQL abordagem com grafos

Fonte: adaptado de Sadalage, Fowler e River (2013)

3 MÉTODO DA PESQUISA

O tipo de pesquisa adotado neste estudo é a pesquisa qualitativa baseada na técnica Delphi. O método de entrevista *Staticized Group* não foi considerado, pois apesar de ser semelhante ao Delphi, limita-se a realização de um único ciclo ou rodada, não buscando o consenso de opiniões dos especialistas de forma sistêmica (HALLOWELL; GAMBATESE, 2010). Como consequência, a aplicação de um *survey* com especialistas, em um painel Delphi, se mostrou a estratégia mais adequada. Além disso, segundo Dickson e Nechis (1984), quando um problema não é definido o suficiente para ser analisado por técnicas analíticas, a técnica Delphi pode ser uma opção viável. Existem poucos estudos relacionados com processo de seleção de SGBD NoSQL, não sendo viável o uso de técnicas analíticas convencionais, justificando assim a aplicação do painel Delphi.

3.1 FASES DA PESQUISA

A pesquisa foi dividida em três fases de modo a atingir os objetivos definidos:

a) primeira fase. A partir de uma revisão da literatura foram identificados os critérios de seleção de SGBDs descritos em trabalhos científicos. Para isso foi feita uma revisão sistemática em bases de dados relacionadas ao tema.

b) segunda fase. A partir dos critérios de seleção de SGBDs obtidos com a revisão da literatura foi aplicada a técnica Delphi. Essa técnica permitiu,

por meio da aplicação de painel com especialistas em banco de dados, obter os principais critérios para avaliação de SGBDs NoSQL.

c) terceira fase. Os critérios de seleção de SGBDs NoSQL obtidos na segunda fase da pesquisa foram ordenados pelo critério de importância, também com a aplicação da técnica Delphi.

3.2 ESCOPO DA PESQUISA

O escopo desta pesquisa abrange organizações privadas. Isto porque, em organizações públicas existem processos de seleção diferenciados, que envolvem inclusive processos de licitação que privilegiam questões econômicas, e que devem ser aderentes a legislação e aos decretos vigentes, diminuindo assim a relevância dos demais critérios de seleção (ZUCCATO JUNIOR, 2014).

Além disso, foram consideradas somente as organizações com aplicações em ambiente Web, pois os critérios se aplicam a este tipo de ambiente, o que justifica a abordagem de SGBDs NoSQL. Isso se deve à relação com *big data*, ou seja, com enorme volume de dados. Observa-se com menos intensidade esta situação em outros tipos de organizações, como por exemplo, em organizações bancárias que necessitam de forte consistência e integridade de dados e nas quais o volume de dados, apesar de grande, não é um gargalo devido aos investimentos realizados por essas organizações (PRADO, 2010).

Por fim, cabe destacar que o presente estudo se restringe a fase de seleção e avaliação de SGBDs NoSQL em organizações privadas com sistemas de informação em ambiente Web.

3.3 TÉCNICA DELPHI

O Delphi é um método que busca obter a opinião coletiva mais assertiva de um grupo de especialistas, aos quais são aplicados questionários individuais com a utilização de um *feedback* controlado durante uma série de ciclos. Neste processo, quando não se atinge um consenso, ao menos ocorre a consolidação e classificação das respostas, ao final da aplicação dos ciclos (MARTINO, 1993).

Quando a técnica Delphi é realizada de forma convencional, costuma produzir um número de, no máximo, quatro rodadas. Um número maior não é aconselhável por restrições de tempo e, principalmente pelo fato de não existirem mudanças de opinião significativas nas rodadas posteriores (KAYO; SECURATO, 1997). Nesta pesquisa o painel Delphi foi aplicado em duas rodadas.

Quanto à forma de aplicação do painel Delphi, foi escolhida para este trabalho a modalidade *ranking form*, por ser possível levantar, de forma ordenada, os critérios de seleção de SGBDs NoSQL mais relevantes.

4 APRESENTAÇÃO DOS RESULTADOS

Os critérios para seleção e avaliação de SGBDs foram extraídos da revisão sistemática e sua fundamentação teórica foi complementada pelo trabalho de Elmasri e Navathe (2005). Esses critérios estão reunidos em quatro grupos e estão apresentados a seguir.

4.1 FORNECEDOR

Após a compilação da lista de SGBDs candidatos, cada fornecedor de software deve ser solicitado para enviar a sua proposta, que pode dar uma visão geral dos recursos que o pacote SGBD oferece (DAVIDSON, 1982). Os fornecedores devem ser solicitados a enviar uma cópia do último relatório anual, no caso de empresas de capital aberto, ou uma descrição da atuação e das atividades da empresa com informações financeiras para avaliar o seu crescimento e estabilidade. Os critérios listados a seguir são usados para avaliar os SGBDs em relação ao fornecedor (DAVIDSON, 1982):

a) reputação. Uma boa reputação está relacionada às respostas obtidas com os pedidos de informação, à forma como foram tratados os problemas relatados e ao apoio técnico fornecido. Para Min (1992), a reputação do fornecedor pode ser considerada como o seu círculo de negócios, o que pode ser uma boa referência para avaliação da credibilidade, boa vontade e confiança no fornecedor.

b) suporte. Representa o apoio dado pelo fornecedor para solucionar problemas com a operação do produto. Alguns desses serviços são gratuitos, outros funcionam vinte e quatro horas por dia incluindo fins de semana.

c) recursos adicionais. Entre os recursos adicionais, três se destacam: (1) fornecimento de bases de testes para avaliação do produto; (2) treinamento do cliente quanto ao uso do SGBD; e (3) disponibilidade de serviço de pós-venda.

d) continuidade. Deve-se verificar o compromisso do fornecedor em manter o produto atualizado e de acordo com o cenário de constante evolução da TI. Alguns SGBDs possuem a tendência de estabilizar conforme amadurecem e podem não ser capazes de suprir as necessidades futuras. As melhorias e modificações são importantes para que o SGBD possa acompanhar a evolução das aplicações. Segundo Min (1992), é desejável escolher um fornecedor estabelecido que esteja em atuação há um longo tempo.

4.2 PRODUTO

Os requisitos operacionais e funcionais que o SGBD deve satisfazer precisam ser documentados. Os requisitos são utilizados para documentar, categorizar e analisar cada pacote candidato. Quando os requisitos priorizados tiverem sido documentados, eles podem ser usados para desenvolver uma lista de critérios para avaliação de SGBDs candidatos. A

seguir destacam-se critérios de avaliação que devem ser considerados em relação ao produto.

a) Desempenho. Critérios relevantes para análise de características de desempenho de um SGBD são: otimização de consultas, execução nativa ou em modo de compatibilidade, capacidade de acessar múltiplas bases e paralelismo (DEWITTS; GRAY, 1992; GHAEMI *et al.*, 2008; LEAVITT, 2010).

b) Administração. Recursos de administração auxiliam os profissionais no gerenciamento e modelagem de dados a realizarem alterações. É desejável que um SGBD forneça bons recursos a fim de agilizar alterações no modelo de dados (DAVIDSON, 1982).

c) Disponibilidade. Precisa ser avaliado se o SGBD deve ser desativado para execução de funções administrativas. Outro aspecto importante é o processamento concorrente, pois uma aplicação pode exigir a capacidade de suportar vários usuários simultâneos (GHAEMI *et al.*, 2008).

d) Flexibilidade. O SGBD deve possuir dicionário de dados integrado com linguagem de consulta e disponibilizar geração dinâmica ou supressão de chaves do usuário e privilégios de acesso de forma flexível. Além disso, deve suportar múltiplos usuários e múltiplos bancos de dados, assim como processamento distribuído (DAVIDSON, 1982).

e) Licença de uso. Pode ser proprietária ou aberta. Na licença proprietária a distribuição do SGBD é controlada e há custos inerentes à contratação do software. Na licença aberta o SGBD pode ser vendido ou disponibilizado gratuitamente (DAVIDSON, 1982).

4.3 ORGANIZAÇÃO

O ambiente de atuação da organização, sua disposição para adotar inovações e assumir riscos podem influenciar os critérios para avaliação e seleção de SGBDs. Segundo Davidson (1982), vários critérios econômicos e organizacionais afetam a escolha de um SGBD, como os descritos a seguir.

a) Custo. Alguns tipos de custos devem ser considerados no processo de avaliação. Entre eles se destacam: custo de aquisição do SGBD, manutenção, aquisição de hardware, criação e conversão do banco de dados, pessoal, treinamento e operacional. Além disso, há custos indiretos que devem ser considerados (BATINI *et al.*, 2009; MIN, 1992).

b) Inovação. A inovação faz parte da filosofia da organização que pode valorizar novas tecnologias ou apenas priorizar a continuidade das aplicações, privilegiando a robustez e a conformidade. Este critério pode ser determinante para a avaliação do uso de um SGBD NoSQL, pois este é um modelo inovador (DAVIDSON, 1982).

c) Perfil de risco. Este critério considera se a organização está disposta a investir em novas tecnologias, mesmo diante dos riscos encontrados durante projetos de mudança (PMI, 2013).

d) Importância da TI. O quanto a organização acredita que a TI seja um fator de adição de valor ao negócio (DAVIDSON, 1982).

e) Profissional de TI. Devem-se considerar também os profissionais da organização envolvidos no processo de avaliação e seleção de um SGBD. Eles representam atores interessados no processo de seleção e podem ter comportamentos diversos (DAVIDSON, 1982).

4.4 QUALIDADE

A qualidade de um pacote de software deve ser avaliada usando um modelo de qualidade definido (KOSCIANSKI; SOARES, 2007). O modelo de qualidade apresentado neste trabalho com o objetivo de identificar critérios de seleção de SGBDs foi o SQuaRE que faz parte da série ISO/IEC 9126 que categoriza os atributos de qualidade de software em seis características, as quais são subdivididas em subcaracterísticas que podem ser medidas por meio de métricas internas e externas.

Os critérios que devem ser considerados em relação à qualidade estão apresentados a seguir e foram resumidos dos trabalhos de Pressman (2009), Loscianski e Soares (2007), NBR ISO/IEC 9126-1 (2003), NBR ISO/IEC 14598-1 (2003), Batini *et al.* (2009) e Davidson (1982).

a) Funcionalidade. Pode-se afirmar que esta característica é idêntica aos “requisitos funcionais”, ou seja, os serviços que se espera que o sistema apresente. Ela pode ser decomposta nas seguintes subcaracterísticas: adequabilidade, acurácia, nível de segurança e *multi-threading*.

b) Manutenibilidade. Está relacionada ao esforço necessário para fazer modificações específicas no software. Basicamente envolve a facilidade de modificação de um produto de software. A manutenibilidade pode ser decomposta nas seguintes subcaracterísticas: modificabilidade, analisabilidade, estabilidade do produto e testabilidade.

c) Usabilidade. Representa o conjunto de atributos que evidenciam o esforço necessário para se utilizar o software, bem como o julgamento individual desse uso. A usabilidade pode ser decomposta nas seguintes subcaracterísticas: operabilidade, compreensibilidade, apreensibilidade e atratividade.

d) Confiabilidade. Está relacionada com um conjunto de atributos que evidenciam a capacidade de manter certo nível de desempenho sob determinadas condições durante um período de tempo estabelecido. A confiabilidade pode ser decomposta nas seguintes subcaracterísticas: tolerância e recuperabilidade.

e) Eficiência. Trata-se do conjunto de atributos que evidenciam o relacionamento entre o nível de desempenho do software e a quantidade de recursos usados, sob condições estabelecidas. A eficiência pode ser decomposta nas seguintes subcaracterísticas: velocidade de operação de um software e utilização de recursos.

f) Portabilidade. Trata-se do conjunto de atributos que evidenciam a capacidade do software de ser transferido de um ambiente para outro, ou seja, de ser utilizado em diferentes plataformas. Pode ser decomposta nas seguintes subcaracterísticas: instabilidade, adaptabilidade, coexistência e substitutibilidade.

5 ANÁLISE DOS RESULTADOS

Esta seção analisa os resultados obtidos com a aplicação do painel Delphi aos critérios identificados na literatura. Primeiro é apresentado o processo de seleção dos especialistas para o painel Delphi, em seguida são descritos os critérios para seleção de SGBDs elencados pelos especialistas, é apresentada a ordem de importância desses critérios, segundo os especialistas, e por fim é analisada a importância dos critérios.

5.1 PROCESSO DE SELEÇÃO DOS ESPECIALISTAS

Os resultados foram analisados junto a um grupo de profissionais que atuam ou têm experiência em SGBD NoSQL, tanto em atividades profissionais em organizações de TI quanto em atividades acadêmicas. Para seleção desse grupo de profissionais foram considerados os seguintes requisitos: (1) anos de experiência em TI; (2) grau de conhecimento e experiência em projetos com banco de dados; (3) participação em processo de seleção de SGBDs NoSQL; (4) formação acadêmica na área de TI; e (5) domínio no desenvolvimento de aplicações ou em projetos suportados por SGBDs NoSQL.

Os profissionais selecionados foram agrupados em três perfis: (1) DAs (*data analyst*) e DBAs (*data base analyst*), que representam os profissionais que atuam diretamente com SGBDs NoSQL, compondo o grupo de administração de dados; (2) desenvolvedores, que representam os profissionais com conhecimento sobre o impacto dos SGBDs NoSQL em projetos de TI e no desenvolvimento de aplicações; (3) demais profissionais de TI, dentre os quais, gerentes, fornecedores de SGBDs e acadêmicos.

Cada perfil tem visões diferentes sobre o processo de seleção de SGBDs, o que permitiu ao grupo ter uma abordagem mais ampla sobre o tema, mesclando experiências e conhecimentos práticos dentro do contexto organizacional. Isso permitiu uma abordagem técnica e gerencial sobre os critérios para seleção de SGBDs NoSQL. A Tabela 1 mostra o resumo do perfil do grupo de 32 profissionais que participaram da avaliação dos critérios de seleção de SGBDs NoSQL.

Tabela 1. Características do grupo de especialistas

Grupo de Especialistas	Experiência TI (anos)		Experiência SGBD		Seleção SGBD		Formação Acadêmica	
	Até 10	Acima de 10	Baixa	Alta	Não	Sim	Até graduação	Pós-graduação
	(18)	(14)	(1)	(31)	(18)	(14)	(28)	(4)
Desenvolvedores	91,7%	8,3%	0,0%	100,0%	91,7%	8,3%	100,0%	0,0%
DAs e DBAs	25,0%	75,0%	0,0%	100,0%	8,3%	91,7%	100,0%	0,0%
Demais profissionais de TI	50,0%	50,0%	12,5%	87,5%	75,0%	25,0%	50,0%	50,0%
Total	56,3%	43,7%	3,1%	96,9%	56,3%	43,7%	87,5%	12,5%

Fonte: elaborada pelos autores

5.2 CRITÉRIOS DE SELEÇÃO PARA SGBD NOSQL

O instrumento utilizado para avaliação dos especialistas foi um questionário *on-line*. A aplicação do instrumento pela Internet foi feita entre os meses de janeiro e março de 2013, utilizando a ferramenta *Survey Monkey*. Ao responder o questionário, cada especialista confirmou, ou não, o grau de importância dos critérios de seleção de SGBD NoSQL identificados na literatura. Todos os critérios identificados na literatura foram avaliados como importantes pelos especialistas e eles ainda acrescentaram mais três critérios, conforme indicado na Figura 5: escalabilidade horizontal, apoio da comunidade e software livre.

Seguem as descrições dos três critérios incluídos pelos especialistas:

a) escalabilidade horizontal. Trata-se da flexibilidade em escalar as aplicações com o aumento na capacidade de processamento por meio do acréscimo no número de máquinas do *cluster* utilizando processamento distribuído.

b) apoio da comunidade. Esse critério verifica a facilidade em se encontrar informações, experiências ou orientações sobre determinado SGBD em comunidades de pesquisa e fóruns na Web.

c) software livre. Este critério se refere à importância da distribuição gratuita do SGBD NoSQL.

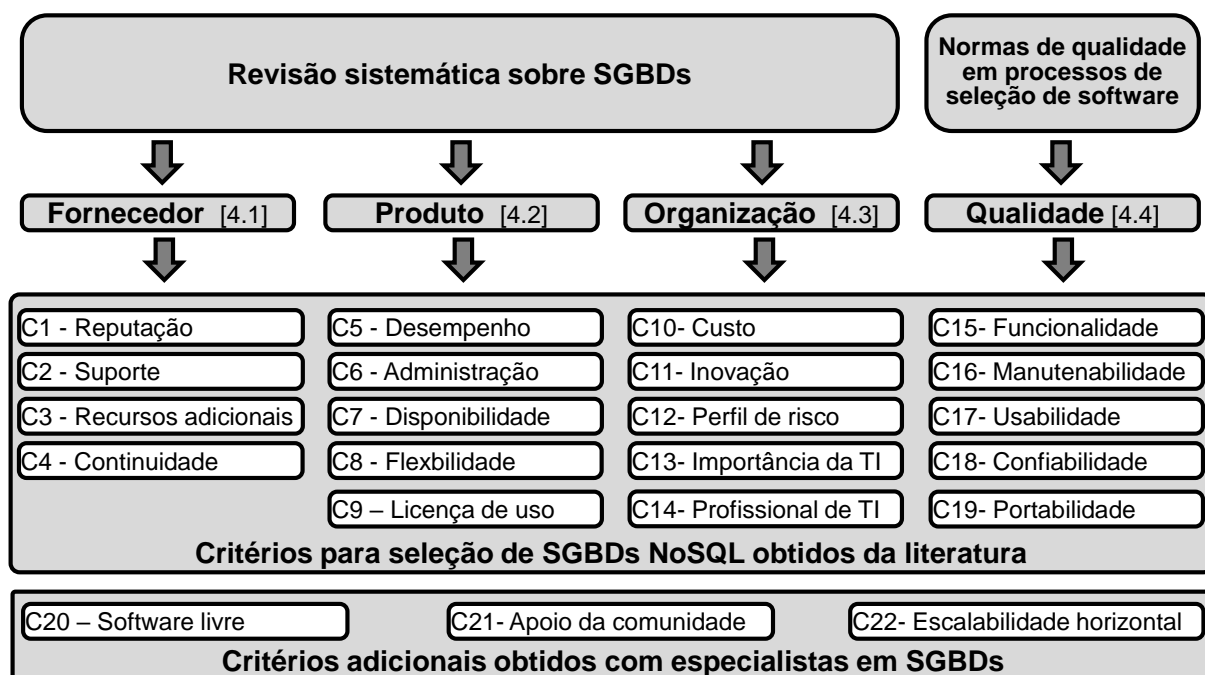


Figura 5. Critérios para seleção de SGBD

Fonte: elaborada pelos autores

5.3 ORDEM DE IMPORTÂNCIA DOS CRITÉRIOS

Foram realizadas duas rodadas no painel Delphi. A primeira rodada apresentou um grau de convergência $W = 0,882$, que segundo Schmidt (1997), é muito alto. Esse resultado teve significância estatística, apresentando um valor de χ^2 (chi-quadrado) igual a 508,01. Segundo Siegel (1981), valores de χ^2 superiores a 59,70 são extremamente significantes. O painel Delphi poderia ter sido encerrado nesta rodada, porém como houve a inclusão de três novos critérios, optou-se por realizar uma nova rodada. A segunda rodada apresentou um grau de convergência $W = 1,052$, superior ao da primeira rodada, mostrando que a concordância entre os especialistas aumentou. Como o resultado apresentou um valor de $\chi^2 = 662,75$, ou seja, extremamente significativo, o painel pode ser encerrado nessa rodada.

O painel Delphi mostrou quais foram os principais critérios para seleção de SGBDs NoSQL, em ordem de importância, sob o ponto de vista de especialistas que atuam em organizações privadas com sistemas de informação em ambiente Web. Os critérios do grupo “Produto” foram os de maior importância segundo os especialistas, uma vez que representam quatro dos nove principais critérios identificados. A Tabela 2 apresenta a classificação dos critérios. O grupo “Qualidade” teve três critérios entre os nove principais, enquanto que no grupo “Fornecedor” foi identificado apenas um critério entre os principais. Os critérios do grupo “Organização” foram os menos importantes uma vez que nenhum deles está entre os nove principais. Por fim, o critério “Apoio da comunidade” (C21) é o único critério que não foi enquadrado em nenhum grupo identificado na literatura.

Analisando a ordem de importância dos critérios de seleção de SGBDs podem-se observar algumas contribuições não esperadas pelo senso comum:

a) Esperava-se que critérios como “Custo” e “Profissional de TI” fossem importantes na escolha de um SGBD NoSQL. Entretanto, os requisitos relacionados à organização tiveram pouca relevância quando comparados aos demais, ou seja, a decisão de escolha de um SGBD NoSQL deve ter predominância técnica apesar da importância do contexto organizacional na adoção de qualquer nova tecnologia.

b) Foi identificado um novo critério não relacionado na literatura e que ficou na lista dos critérios mais relevantes para seleção de SGBDs NoSQL: “Apoio da comunidade” (C21). Para a escolha de um SGBD NoSQL, os especialistas consideraram importante que ele tenha bom suporte da comunidade, como por exemplo, a existência de fóruns especializados no tema. O grupo de “Desenvolvedores” foi o que pontuou este critério com mais intensidade. Além disso, analisando apenas os gerentes pertencentes ao grupo Demais Profissionais de TI, este critério ficou na sétima posição na ordem de importância, mostrando que atualmente os gerentes consideram importante o apoio da comunidade, em oposição a uma postura anterior em que somente o apoio dos fornecedores era considerado relevante.

Tabela 2. Classificação dos critérios por ordem de importância

Critérios de seleção de SGBDs	Classificação por grupos			
	Todos os especialistas	Desenvolvedores	DBAs e DAs	Demais profissionais de TI
Produto				
C5 – Desempenho	1º	2º	1º	2º
C7 – Disponibilidade	3º	4º	3º	3º
C22 – Escalabilidade horizontal	4º	3º	8º	4º
C8 – Flexibilidade	7º	6º	4º	9º
Qualidade				
C12 – Confiabilidade	2º	1º	2º	1º
C11 – Manutenibilidade	5º	5º	5º	5º
C13 – Funcionalidade	6º	8º	6º	6º
Fornecedor				
C4 – Capacidade de investimento	8º	9º	7º	7º
Comunidade				
C21 – Apoio da comunidade	9º	7º	9º	11º

Fonte: elaborada pelos autores

c) O critério “Flexibilidade” (C8) teve melhor classificação para o grupo DBAs e DAs, ocupando a quarta posição neste grupo. Resultado interessante, pois para o senso comum esse critério deveria ter maior impacto para o grupo de “Desenvolvedores”. Isto porque, os SGBDs que apresentam problemas de flexibilidade impactam diretamente nas aplicações, dificultando o processo de alteração no modelo de dados e tornando mais longo o processo de desenvolvimento.

d) Esperava-se que o grupo “Demais profissionais de TI”, composto principalmente por gestores, privilegiasse critérios ligados a aspectos organizacionais de TI no processo de seleção de SGBDs NoSQL. Entretanto, este grupo manteve comportamento parecido com os demais grupos, dando maior importância para critérios técnicos e relacionados especificamente ao SGBD, como “Confiabilidade” e “Desempenho”.

Analisando os dois critérios mais importantes (desempenho e confiabilidade), pode-se, num primeiro momento, identificar um conflito, pois alguns SGBDs NoSQL abrem mão de consistência e integridade, que conferem confiabilidade às aplicações, para a obtenção de desempenho. É importante esclarecer que o critério confiabilidade foi entendido dentro do escopo deste estudo, que são organizações privadas com sistemas de informação em ambiente Web, e nas quais a confiabilidade se relaciona também à continuidade do funcionamento correto em situações inesperadas e de erro. Dessa forma, não há conflito entre o critério de desempenho e o de confiabilidade. Corrobora essa argumentação a existência de SGBDs NoSQL oferecidos no mercado brasileiro de TI, que

garantem o mesmo nível de desempenho em situações inesperadas ou eventuais problemas, garantindo também a confiabilidade.

5.4 ANÁLISE DA IMPORTÂNCIA DOS GRUPOS DE CRITÉRIOS

Após a análise da importância dos critérios fez-se uma análise da importância dos grupos de critérios. Compararam-se as médias das pontuações obtidas para cada grupo utilizando a técnica estatística de análise de variância (ANOVA), conforme mostra a Tabela 3. Os grupos com pontuação mais alta foram respectivamente, “Produto”, “Qualidade”, “Comunidade” e “Fornecedor”. Os grupos Produto e Qualidade têm pontuação média acima dos grupos “Comunidade” e “Fornecedor” com nível de significância estatística de 5%. Entretanto, não há diferença estatística significativa entre os grupos Produto e Qualidade, e entre os grupos Fornecedor e Comunidade. Isto mostra que os painelistas consideraram os critérios associados ao produto e à qualidade como sendo os mais importantes e tendo a mesma importância entre si.

Tabela 3. Comparação entre os grupos de critérios de seleção

Critérios de seleção de SGBDs	N	Média	Diferença entre as médias (sig.*)			
			Produto	Qualidade	Fornecedor	Comunidade
Produto	30	4.267		0.405	0.002	0.000
Qualidade	30	4.134	0.405		0.023	0.004
Fornecedor	30	3.767	0.002	0.023		0.532
Comunidade	30	3.958	0.000	0.004	0.532	

Legenda: * ANOVA. Valores menores que 0,05 têm nível de significância estatística de 5%.

Fonte: elaborada pelos autores

6 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi identificar e analisar os principais critérios que devem ser usados para avaliação de um SGBD NoSQL, com base na opinião de especialistas em banco de dados. Para atingir os objetivos propostos, efetuou-se uma pesquisa descritiva e qualitativa utilizando a técnica Delphi com a colaboração de 32 especialistas. Cabe destacar que a presente pesquisa apresenta limitações. Entre elas, destacam-se: as limitações relacionadas ao método, pois o uso do painel Delphi não permite generalização dos resultados e pode ocorrer excesso de influência por parte de um ou vários painelistas. O método de pesquisa também pode ter gerado viés, uma vez que a análise do painel Delphi, dividida por grupos de especialistas, difere da abordagem tradicional que analisa os especialistas como um único grupo. Apesar das limitações apresentadas, os resultados obtidos são satisfatórios face aos objetivos propostos.

Verificou-se que a literatura possui pouco conteúdo explorando a questão do processo de seleção de SGBDs. Nesse sentido, o painel Delphi contribuiu para a consolidação de 22 critérios que podem ser utilizados para avaliação e seleção de SGBDs NoSQL.

Os critérios apresentados estão associados principalmente às características de produto, ou seja, às características do SGBD em si. O principal critério escolhido pelos especialistas (C5 - Desempenho) é corroborado pelo trabalho de Leavitt (2010) que destaca o desempenho como principal característica dos SGBDs NoSQL. Pode-se concluir que os especialistas consideram como principais critérios para seleção de SGBDs NoSQL as características técnicas do SGBD.

A abrangência do tema *big data* e o massivo volume de dados exigido nas aplicações Web possibilitam o desenvolvimento de diversos estudos na área de Sistema de Informação. Nesse contexto, recomenda-se a elaboração de novas pesquisas, a partir dos resultados deste trabalho. Entre as várias alternativas, destacam-se: (1) pesquisas relacionadas ao papel da comunidade, que pode facilitar o processo de implantação de uma nova tecnologia com a disponibilização de informações compartilhadas entre os profissionais; e (2) pesquisas quantitativas com uso de amostras que permitam generalizar resultados.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR ISO/IEC 9126-1: engenharia de software – qualidade de produto – parte 1: modelo de qualidade. Rio de Janeiro - RJ, 2003.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR ISO/IEC 14598-1: tecnologia de informação – avaliação de produto de software – parte 1: visão geral. Rio de Janeiro - RJ, 2003.

BATINI, C.; CAPPIELLO, C.; FRANICALANCI, C.; MAURINO, A. Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, v. 41, n. 3, p. 1-52, 2009. <http://dx.doi.org/10.1145/1541880.1541883>

CHANG, F.; DEAN, J.; GHEMAWAT, S. B. A distributed storage system for structured data. *ACM Transactions on Computer Systems (TOCS)*, v. 26, n. 2, article no. 4, June, 2008.

DAVIDSON, E. Evaluating database management systems. In: *Proceedings of the June 7-10, National Computer Conference (AFIPS '82)*, ACM, New York, NY, USA, 1982. <http://dx.doi.org/10.1145/1500774.1500855>

DEAN, J.; GHEMAWAT, S. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, v. 51, n. 1, p. 107-113, 2008. <http://dx.doi.org/10.1145/1327452.1327492>

DEWITT, D.; GRAY, J. Parallel database systems: the future of high performance database systems. *Communications of the ACM*, v. 36, n. 6, June, 1992. <http://dx.doi.org/10.1145/129888.129894>

DICKSON, G. W.; NECHIS, M. Key information system issues for the 1980's. *MIS Quartely*, v. 8, n. 3, p. 135-159, 1984. <http://dx.doi.org/10.2307/248662>

ELMASRI, R.; NAVATHE, S. B. *Sistemas de banco de dados*. Pearson Education, Brasil, 2005.

- GHAEMI, R.; FARD, A. M.; TABATABAEE, H.; SADEGHIZADEH, M. Evolutionary query optimization for heterogeneous distributed database systems. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, v. 2, n. 7, 2008.
- GHEMAWAT, S.; GOBIOFF, H.; LEUNG, S. The Google file system. *SIGOPS Operating Systems Review*, v. 37, p. 29-43, 2003. <http://dx.doi.org/10.1145/1165389.945450>
- HALLOWELL, M. R.; GAMBATESE, J. A. Qualitative research: application of the Delphi method to CEM research. *Journal of Construction Engineering and Management*, v. 136, n. 1, p. 99-107, 2010. [http://dx.doi.org/10.1061/\(ASCE\)CO.1943-7862.0000137](http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000137)
- HARRISON, Guy. *NoSQL and document-oriented databases*. Database Trends and Applications, 2010. Disponível em <<http://www.dbta.com/Columns/Notes-on-NoSQL/NoSQL-and-Document-OrientedDatabases-72035.aspx>>. Acesso em: 03 março 2014.
- KAYO, E. K.; SECURATO, J. R. Método Delphi: fundamentos, críticas e vieses. *Cadernos de Pesquisa em Administração*, n. 1, p. 51-61, 1997.
- KOSCIANSKI, A.; SOARES, M. Qualidade de software: aprenda as metodologias e técnicas mais modernas para o desenvolvimento de software, 2. ed. São Paulo: Novatec Editora, 2007.
- LEAVITT, N. Will NoSQL Databases Live Up to Their Promise? *IEEE Computer*, v. 43, n. 2, pp. 12-14, 2010. <http://dx.doi.org/10.1109/MC.2010.58>
- LÓSCIO, B. F.; OLIVEIRA, H. R.; PONTES, J. C. S. NoSQL no desenvolvimento de aplicações Web colaborativas. In: VIII *Simpósio Brasileiro de Sistemas Colaborativos*, Brasil, 2011.
- MARTINO, J. P. *Technological forecasting for decision making*. McGraw-Hill Engineering and Technology Management Series. 3rd ed, 1993.
- MIN, H. Selection of software: the analytic hierarchy process. *International Journal of Physical Distribution & Logistics Management*, v. 22, n. 1, p. 42-52, 1992. <http://dx.doi.org/10.1108/09600039210010388>
- PMI. *A guide to the project management body of knowledge (PMBOK Guide)*, Newton Square: Project Management Institute, v. 5, 2013.
- POKORNY, J. NoSQL databases. A step to database scalability in web environment. In: *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (iiWAS '11)*. ACM, New York, NY, USA, pp. 278-283, 2011.
- PRADO, E. P. V. Sites de Internet banking: uma avaliação da qualidade baseada no modelo WEBQUAL. *Gestão & Regionalidade*, v. 26, n. 77, p. 63-74, 2010. <http://dx.doi.org/10.13037/gr.vol26n77.249>
- PRESSMAN, R. S. *Software engineering: a practitioner's approach*, 7th ed. New York: McGraw-Hill, 2009.

RAHIEN, A. That No SQL Thing: Column (Family) Database, 2010. Disponível em: <<http://ayende.com/blog/4500/that-no-sql-thing-column-family-databases>>. Acesso em: 03 março 2014.

REDMOND, E.; WILSON, J. R. *Seven databases in seven weeks: a guide to modern databases and the NoSQL movement*. Dallas: Pragmatic Programmers LLC., 2012.

SADALAGE, P. J.; FOWLER, M.; RIVER, U. S. *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Pearson Education, Brasil, 2013.

SCHMIDT, R. C. Managing Delphi Surveys Using Nonparametric Statistical Techniques. *Decision Sciences*, v. 28, n. 3, p. 763-774, 1997. <http://dx.doi.org/10.1111/j.1540-5915.1997.tb01330.x>

SIEGEL, S. *Estatística não-paramétrica: para ciências do comportamento*. 1981.

TIWARI, S. Professional NoSQL, Indianapolis: John Wiley & Sons, Inc, 2011.

YE, Z.; LI, S. A request skew aware heterogeneous distributed storage system based on Cassandra. In *Proceedings of the International Conference on Computer and Management (CAMAN '11)*, p. 1-5, maio 2011. <http://dx.doi.org/10.1109/caman.2011.5778745>

ZUCCATO JUNIOR, F. *Gerenciamento das aquisições em projetos*, 1. ed. Rio de Janeiro: Elsevier, 2014.